idTarget: a web server for identifying protein targets of small chemical molecules with robust scoring functions and a divide-and-conquer docking approach

Jui-Chih Wang[1], Pei-Ying Chu[2], Chung-Ming Chen[1] and Jung-Hsin Lin[2,3,4,*]

[1]Institute of Biomedical Engineering, National Taiwan University, [2]Division of Mechanics, Research Center for Applied Sciences, Academia Sinica, [3]School of Pharmacy, National Taiwan University and [4]Institute of Biomedical Science, Academia Sinica, Taipei, Taiwan

ABSTRACT

Identification of possible protein targets of small chemical molecules is an important step for unravelling their underlying causes of actions at the molecular level. To this end, we construct a web server, idTarget, which can predict possible binding targets of a small chemical molecule via a divide-and-conquer docking approach, in combination with our recently developed scoring functions based on robust regression analysis and quantum chemical charge models. Affinity profiles of the protein targets are used to provide the confidence levels of prediction. The divide-and-conquer docking approach uses adaptively constructed small overlapping grids to constrain the searching space, thereby achieving better docking efficiency. Unlike previous approaches that screen against a specific class of targets or a limited number of targets, idTarget screen against nearly all protein structures deposited in the Protein Data Bank (PDB). We show that idTarget is able to reproduce known off-targets of drugs or drug-like compounds, and the suggested new targets could be prioritized for further investigation. idTarget is freely available as a web-based server at http://idtarget.rcas.sinica.edu.tw.

INTRODUCTION

Identification of targets of small chemical molecules is essential for unravelling the underlying molecular causes of actions. Often, natural products, i.e. compounds discovered from plants, animals, marine lives or other living organism, exhibit useful pharmaceutical effects, e.g. anti-inflammatory, anti-cancer and anti-viral effects, yet their molecular mechanisms remain elusive. On the other hand, many drugs are known to be accompanied with unpleasant adverse effects, but the molecular targets of such effects are largely unknown. On the contrary, there are also some old drugs whose additional beneficiary effects are discovered only recently. For example, the epigenetic mechanism of the anticancer effect of cholesterol-lowering drugs, statins, was uncovered rather recently (1).

Conventional virtual screening of chemical libraries has been used widely to search for new leads in drug development for a protein target (2). As the deposited structures of biomolecules in the Protein Data Bank (PDB) increase substantially in the past decades, searching for the targets of a given drug or small compounds (also known as inverse screening, target fishing, off-target prediction, etc.) has become a useful approach (3–7).

One of the major hurdles for target identification is the effectiveness of scoring functions (7,8). To evaluate the binding affinity of the small ligand and a protein target, an accurate yet generally applicable scoring function is essential. We recently developed three robust scoring functions, AutoDock4^{RRP}, AutoDock4^{RAP} and AutoDock4^{RGG} (9) based on the energetic terms and the formulation of AutoDock4 (10). These scoring functions report the binding free energy in the experimental scale, which allows direct comparison among different protein–ligand systems. Two of these three robust scoring functions were constructed using atomic charges from quantum chemical calculations, namely, RESP (11) and AM1-BCC (12), and the robust regression analysis (13) was employed to mitigate the influence of outliers for the calibration of the scoring functions. These robust AutoDock4 scoring functions have been benchmarked for their capability in binding affinity prediction and binding pose prediction (9). For the assessment of binding affinity prediction with a large external set of

1427 complexes from PDBbind v2009, AutoDock4^{RAP} obtained root-mean-square errors of 2.176 kcal/mol, while the size of the training set is only 147. Benchmarked by using two decoy sets (14,15), the robust AutoDock4 scoring functions outperformed most of other scoring functions for the binding pose prediction (9).

Here, we utilize an efficient docking approach to screen the protein targets. Evaluation of potential targets is carried out by using the AutoDock4 robust scoring functions and the affinity profile analysis to enhance the confidence level of prediction.

MATERIALS AND METHODS

Docking and scoring

The search engine of idTarget web server is MEDock (16), which generates initial docking poses of the small ligand. The global search algorithm used in MEDock has also been tested recently by random mathematical functions simulating rugged free energy landscapes with different dimensionalities (17). It was shown that this global search algorithm maintained very high searching efficiency even at the dimensionality of 30 (17), which should be sufficient for applying to most protein–ligand systems. It was also shown that the traditional genetic algorithm failed to deliver good searching efficiency as long as the dimensionality is beyond 20 (17).

Since the binding sites are generally not known, the search space (i.e. the docking grid), should cover the entire protein. However, the searching efficiency also depends on the size of the search space. To achieve better searching efficiency, as well as to overcome the memory limitation when docking to very large proteins, a divide-and-conquer docking approach, similar to the partial box approach in BDT (18), was adopted for efficient blind docking. It should be emphasized that the entire receptor surface is searched, instead of only a part of the surface or just the active site. This also allows the search for possible allosteric binding sites. The divide-and-conquer docking approach can easily take the advantage of the parallel computing facilities and reduces the searching time by limiting the size of grids. The large box that covers the entire protein surface is subsequently divided into smaller boxes, where the size of the smaller boxes is dynamically determined according to the size of query ligand. If a grid box is far away from the receptor, namely, no atom is within 1.42 times the length of the grid box to the center, it will be eliminated to reduce the computational cost. The pre-calculated energetic grid maps for the protein targets are stored in the idTarget web server, and therefore the time for grid construction can be saved.

To analyze the results of target screening, the affinity profile of each binding site is constructed. Our affinity profile analysis was inspired by the structure-based maximal affinity model recently proposed (19). Here, instead of maximum affinity, the affinity profile of a given pocket is modelled by a Gaussian function, whose width is determined by the range of the predicted binding affinities of the complexes of different ligands in the same binding pocket of the same protein. A Z-score of ligand j to a given protein pocket i is calculated based on above profile using the following equation:

$$Z_{ij} = \frac{E_{ij} - E_i}{sd_i}$$

$$E_i = \sum_{k=1}^{N_c} E_k^c / N_c$$

$$sd_i = \sqrt{\sum_{k=1}^{N_c} \left(E_k^c - E_i\right)^2 / N_c - 1}$$

where E_{ij} is the dock energy of ligand j docking to the protein pocket i. E^c represents the energy directly evaluated with the crystal pose. E_i and sd_i are the center and the width of the affinity profile of protein pocket i, respectively. A large negative Z value signifies an important target of the query compound.

The contraction-and-expansion strategy

Since direct screening for a large set of targets is computational intensive, a contraction-and-expansion strategy is adopted to perform target screening efficiently. In the contraction stage, the set of representative pockets for the protein–ligand complexes in the PDB, 3046 mean points of scPDB (20,21), were clustered by CD-Hit (22) under the cutoff of 40% sequence identity. After clustering of protein sequences in the PDB, a contracted list of targets (2091 targets) was obtained. The targets screening starts from this contracted list, subsequently, half of targets with lower docking energy are kept. In the expansion stage, targets that are homologous or contain similar binding sites to the top representatives are also selected for screening. The idTarget server is periodically updated to include new protein structures from PDB. The pockets of a new PDB entry will be compared with the set of representative pockets by SiteEngine (23) and subsequently assigned to the most similar pocket group.

In addition to the contraction-and-expansion strategy, idTarget provides two modes for searching binding poses. In the 'scanning mode', conventional docking jobs are performed for each protein structure. In the 'fast mode', the ligand will first be mapped to the binding sites of the homologous proteins by superposition of homologous protein structures with the pre-calculated structural alignment information by using CE (24), and then its binding poses are optimized by adaptive local searches to remove too close contacts with protein atoms. To extend the coverage of PDB structures, idTarget also performs a ligand similarity search to find structurally similar co-crystallized ligands. The Tanimoto coefficients calculated by OpenBabel 2.3.0 (25) are used to rank the results of the similarity search by matching molecular fingerprints.

INPUT

The input window of the idTarget web server requires an input ligand file for the target screening. Users can upload a ligand file with the popular molecular file formats (pdb,

mol2, pdbqt or cif). Internally, idTarget uses the 'pdbqt' file format that contains the information of atomic charges and atom types. Users can select the Gasteiger, AM1-BCC or RESP model to calculate the atomic charges for the uploaded ligand. Otherwise, the default atomic charges will be calculated by the AM1-BCC charge model. idTarget accepts an ligand file with predetermined protonated state by checking the option, 'Yes, no further addition of polar hydrogens is needed', or the protonation state of the ligand will be assigned by the idTarget server. The idTarget web server also provides a graphical user interface using the Marvin Sketch applet for directly drawing a molecule on the web interface. As the default setting, the larger scale screening is performed with the idTarget data set. Users can also query a small set of proteins with known structures by editing a PDB ID list.

OUTPUT

Tracking the status of a job and retrieving the results can be achieved by input the Task ID. If the e-mail address is provided, the notification will be sent to users as soon as the calculation is finished. Although it usually takes hours to finish a job, depending on the loading of the server, users can always check the results while some targets have been screened. During the calculation of a job, the screening results are sorted immediately for showing the potential targets in the higher ranks. As depicted in Figure 1, two hierarchical tables and a molecular graphic visualizer are shown in the result page. The two-layer table (grouped by protein name) lists all the results sorted by dock energies. The PDB entries with the same target name will be grouped, and one can click on the button to expand/collapse the next layer or all results at once. The homologous proteins are grouped in the three-layer table (grouped by the binding pocket), where the representative one is shown in the first layer and the proteins within the same binding pocket are shown in the second layer. All the PDB entries with the same target name are collected in the third layer. Users can also use selected ranges of predicted binding free energy and Z-score as the filtering criteria to obtain a more condensed target list.

WEB SERVER

The web interface of idTarget was written in PHP (5.1.6). Marvin Sketch applet (http://www.chemaxon.com/products/marvin/marvinsketch/) was embedded in the idTarget job submission page so that users can draw their interested ligand online. In the result page, users can inspect the binding poses and manipulate the resulting 3D structures interactively via the embedded OpenAstexViewer applet (http://openastexviewer.net/web/).

EXAMPLES

To demonstrate the usage of the idTarget web server, three examples are presented here. The first example is the HIV-1 protease inhibitor, darunavir (DRV), which has been approved by FDA on 2006 and has been validated the inhibition against multidrug-resistant viruses (26). The second example is one of indirubin derivatives, 6-bromo-indirubin-3′ oxime (6BIO), which has been experimentally validated as an inhibitor of several kinds of protein kinases (27). The third example is an inhibitor of histone deacetylase 2 (HDAC2), namely, *N*-(4-aminobiphenyl-3-yl)-benzamide (dubbed 'LLX' from its PDB component ID). The target screening results of these three examples are archived in the web server with the Task IDs DRV_screening, 6BIO_screening and LLX_screening, respectively.

DRV

Mutations of HIV-1 protease (PR) residues usually reduce susceptibility to the PR inhibitors but not significantly to darunavir. There have been many structural and kinetic analyses of darunavir against several drug-resistant mutants (26,28–30). Table 1 gives the comparison of predicted and experimental binding affinities for the wild-type and mutant HIV-1 PR. These results indicate the comparable predicted binding affinity to the experimental data. Although the scoring function was not trained with these PR-inhibitor complexes, it still can discriminate the fine difference of affinity against mutants, except for the case of 3CYW. It should be noted that in the docking, we have excluded a well-known flap water molecule, which forms a hydrogen bond network between the flaps of PR and ligand. If that flap water was included in docking of DRV, the predicted free energy will be lowered by ∼1 kcal/mol. This flap water was removed because in general it may cause steric clashes for most ligands. Figure 2 depicts the binding pose of DRV against the wild-type PR with the rmsd of 0.85 Å compared to the crystal pose.

6BIO

This example shows that not only the well-known kinase targets (CDK2, CDK5 and GSK-3β) could be identified by idTarget, but also the kinase PDK1 (27) and other proteins. More detailed processes are described in 'Example 1' in our idTarget web server. The predicted best targets are as follows: CDK2 (PDB ID: 1FVV, predicted K_i: 51.7 nM); CDK5 (PDB ID: 1UNH, predicted K_i: 71.2 nM); GSK-3β (PDB ID: 1Q41, predicted K_i: 63.3 nM); PDK1 (PDB ID: 2PE1, predicted K_i: 96.5 nM). Figure 3 depicts the best docking pose of 6-bromo-indirubin-3′ oxime (6BIO) in PDK1 that confirms the proposed pose in Zahler *et al.* (27).

To compare directly with the results of Zahler *et al.* (27), we performed the screening with the same set of 5821 PDB entries by using the 'User-edited list' feature. The predicted binding free energy was used to rank the screening results, and the Z-score (<0) was the filter. For the known protein kinase targets (CDK2, CDK5 and GSK-3β), the top 1% and top 5% enrichment factors in our screening results were 6.45 and 11.94, respectively, while the top 1% and top 5% enrichment factors calculated from the results of Zahler *et al.* were 19.35
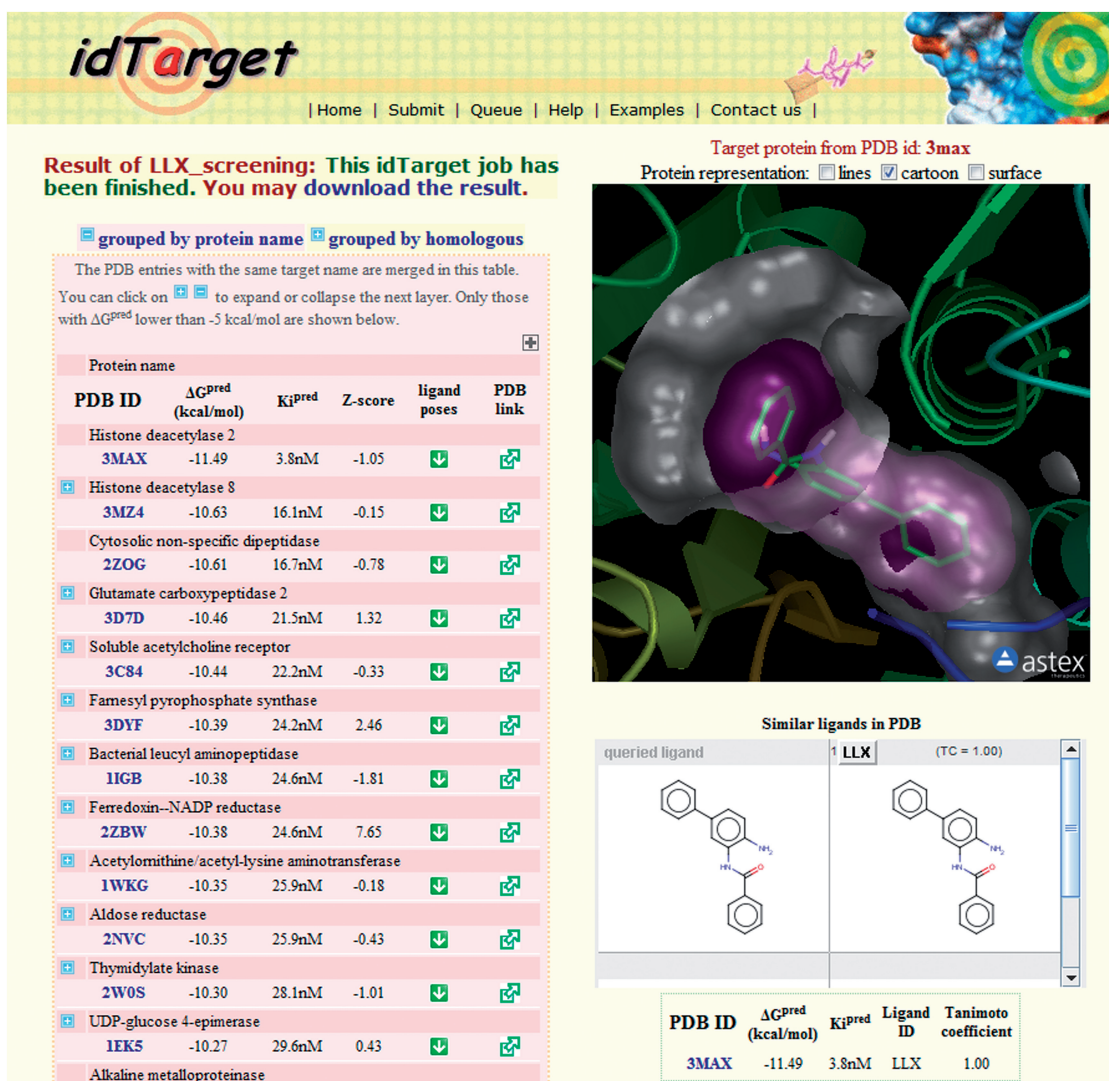
Figure 1. An example of target screening for the HDAC2 inhibitor (LLX). The left panel gives two kinds of representations: grouped by protein name and grouped by homology. The right-hand side contains a viewer for visualizing the docking poses online and a panel showing 2D-similar ligands in PDB.

Table 1. Darunavir screening results for wild-type and mutant HIV-1 PR

PDBID	$\Delta G^{pred.}$	$Ki^{pred.}$(nM)	Z-score	mutant	$Ki^{exp.}$ (nM)
3CYW	−10.99	8.8	0.26	G48V	17[a]
2F81	−10.82	11.7	0.32	L90M	0.03[b]
2IEN	−10.79	12.3	0.33	wild	0.22–1.0[a,b,c]
2IDW	−10.65	15.6	0.38	V82A	0.8–1.3[b,c]
2IEO	−10.55	18.5	0.41	I84V	3.2[b,c]
1T3R	−10.53	19.1	0.42	wild	0.06[d]
2HS1	−10.47	21.1	0.44	V32I	3.3[e]
2F80	−10.4	23.8	0.46	D30N	6.6[b]
2F8G	−10.38	24.6	0.47	I50V	2.0–18[a,b]
3D1Z	−10.38	24.6	0.47	I54M	1.6[a]
3D20	−10.38	24.6	0.47	I54V	5[a]

[a]Experimental data from *J. Mol. Biol.* (2008) **381**, 102–115.
[b]Experimental data from *J. Med. Chem.* (2006) **49**, 1379–1387.
[c]Experimental data from *J. Mol. Biol.* (2004) **338**, 341–352.
[d]Experimental data from *J. Med. Chem.* (2005) **48**, 1813–1822.
[e]Experimental data from *J. Mol. Biol.* (2006) **363**, 161–173.

and 7.42, respectively. It should be noted that protein kinases are the only targets of interest in the work of Zahler *et al*. However, we found that two proteins, matrix metalloproteinase and glycogen phosphorylase, constitute the major part of the top 1% of our results. Interestingly, we found that a recent experimental work (31) showed that an indirubin derivative inhibits rabbit muscle glycogen phosphorylase *b*. The enrichment factor for glycogen phosphorylase (PYGM) in the top 1% was 18.18 in our results, but none of PYGM was identified in the top 5% of the results of Zahler *et al*. We further performed the ROC curve analysis (32) to assess of the performance of the prediction. Figure 4 shows the ROC curves and numerical values of the areas under the curve (AUC), where CDK2, CDK5 and GSK-3β are considered as known targets of 6BIO (in Figure 4a), and these three kinases plus PYGM are considered as known targets (in Figure 4b), respectively. Supplementary Figures S1–S4

show the ROC curves with different selections of known targets and *Z*-score filters. The more the stringent of the *Z*-score filter, the higher the AUC was obtained, which could be due to the nature of conformational selectivity,
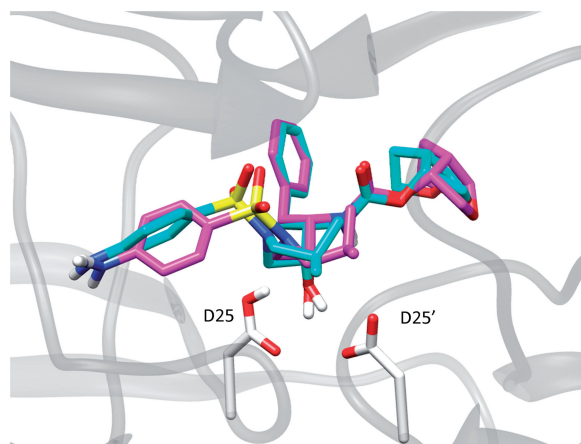

Figure 2. The docking pose of darunavir (in magenta) compared to the x-ray pose (in cyan) in the wild-type protease (PDB ID: 2IEN) gives a RMSD of 0.846 Å.
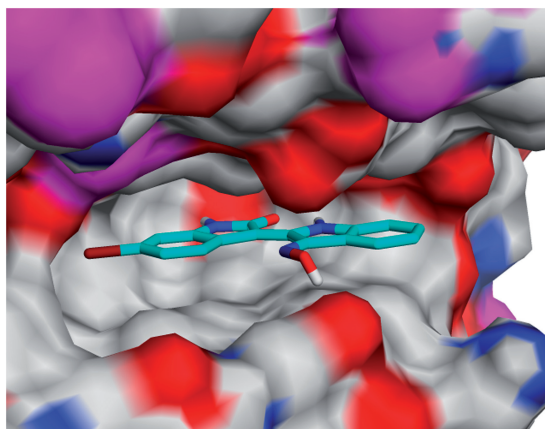

Figure 3. The best docking pose of 6-bromo-indirubin-3' oxime (6BIO) in PDK1.

the general applicability and the universal physical scale of our robust scoring function (9). These analyses also indicate that our predictions are comparable with the prediction of Zahler *et al.*

LLX

N-(4-aminobiphenyl-3-yl)-benzamide is a compound recently designed and synthesized to optimize the fitting of the HDAC2 binding pocket (33). It has been long recognized that there is a 14-Å long internal cavity, also called 'foot pocket', adjacent to the catalytic zinc ion of HDAC. This foot pocket connects to a lipophilic tube with an angle of about 110°, and this compound sits at the kink between the foot pocket and the tube. As shown in Figure 1, the screenshot of the result page given by idTarget, the compound LLX recognized its co-crystalized protein structure (PDB ID: 3MAX) with the highest affinity.

DISCUSSION

Although darunavir is a relatively new drug, its adverse effects and interactions with other drugs have been reported (34,35). It may be possible to further investigate possible molecular causes of actions of these adverse effects and drug–drug interactions by scrutinizing the list of proteins identified by idTarget. Our prediction for possible targets of 6BIO is generally consistent with the previous work. (27). It is interesting to note that the top prediction of LLX is 3MAX, which is the HDAC2 structure crystallized by the same group that designs and synthesizes this *N*-(4-aminobiphenyl-3-yl)-benzamide. A possible explanation could be that this compound sits at the kink between the foot pocket and the lipophilic tube, which altogether forms a very curved channel structure and may be rather unique among protein structures.

Drugs can bind to a myriad of particles in the blood, such as red blood cell, leukocytes, platelets, as well as to some proteins, e.g. serum albumin, lipoproteins, α_1-acid glycoproteins, α, β, γ-globulins, etc. (36). Usually, the
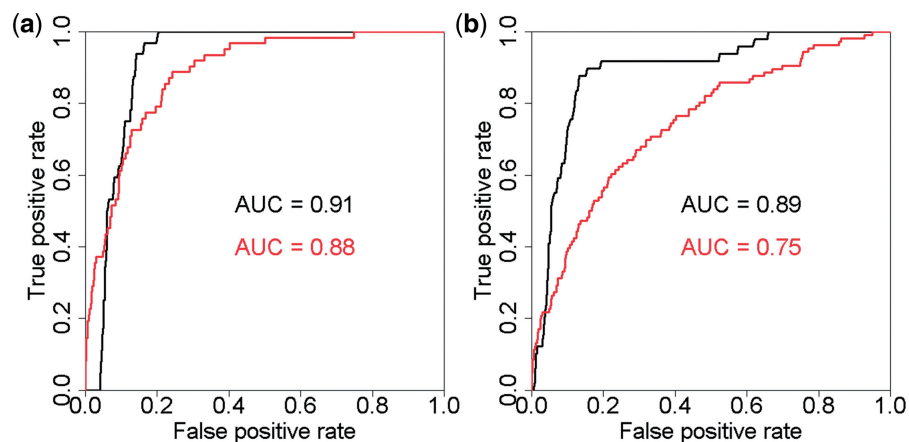

Figure 4. The red curve is obtained from the prediction of Zahler *et al.*, while the black curve is obtained from the prediction of idTarget with *Z*-score < 0 set as the selection filter (1161 protein targets selected). (**a**) CDK2, CDK5 and GSK-3β are considered as known targets and others are considered as decoys. (**b**) CDK2, CDK5, GSK-3β and PYGM are considered as known targets and others are considered as decoys.

metabolites of a drug (or a small molecule) can also bind to other human proteins to exert unwanted effects, or even bind to the original target protein and deliver similar actions. In view of this, target identification should be employed in tandem with metabolite prediction algorithms, e.g. MetaSite (37). Many drugs, e.g. vinblastine, verapamil, taxol, etc., are substrates of P-glycoprotein, whose low resolution structure has been available recently (38), and higher resolution structures should be on the way (G. Chang, personal communication). Due to the tremendous cost of large-scale protein binding assays, the comprehensive list of the proteins that a given compound binds is generally not available. One should also bear in mind that proteins are localized in compartmented organelles, and a given compound may not be able to be distributed to a specific compartment where a certain protein exists, and therefore the prediction of target identification should also be guided by the knowledge of protein subcellular locations to remove the false positives if only the physiological scenario is of interest.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figures 1–4.

REFERENCES

1. Lin,Y.C., Lin,J.H., Chou,C.W., Chang,Y.F., Yeh,S.H. and Chen,C.C. (2008) Statins increase p21 through inhibition of histone deacetylase activity and release of promoter-associated HDAC1/2. *Cancer Res.*, **68**, 2375–2383.
2. Shoichet,B.K. (2004) Virtual screening of chemical libraries. *Nature*, **432**, 862–865.
3. Chen,Y.Z. and Zhi,D.G. (2001) Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins-Struct. Funct. Genet.*, **43**, 217–226.
4. Keiser,M.J., Setola,V., Irwin,J.J., Laggner,C., Abbas,A.I., Hufeisen,S.J., Jensen,N.H., Kuijer,M.B., Matos,R.C., Tran,T.B. *et al.* (2009) Predicting new molecular targets for known drugs. *Nature*, **462**, 175–U148.
5. Yang,L., Luo,H., Chen,J., Xing,Q.H. and He,L. (2009) SePreSA: a server for the prediction of populations susceptible to serious adverse drug reactions implementing the methodology of a

6. chemical–protein interactome. *Nucleic Acids Res.*, **37**, W406–W412.
6. Wale,N. and Karypis,G. (2009) Target fishing for chemical compounds using target-ligand activity data and ranking based methods. *J. Chem. Inf. Model.*, **49**, 2190–2201.
7. Kellenberger,E., Foata,N. and Rognan,D. (2008) Ranking targets in structure-based virtual screening of three-dimensional protein libraries: methods and problems. *J. Chem. Inf. Model.*, **48**, 1014–1025.
8. Xie,L. and Bourne,P.E. (2011) Structure-based systems biology for analyzing off-target binding. *Curr. Opin. Struct. Biol.*, **21**, 189–199.
9. Wang,J.C., Lin,J.H., Chen,C.M., Perryman,A.L. and Olson,A.J. (2011) Robust scoring functions for protein–ligand interactions with quantum chemical charge models. *J. Chem. Inf. Model.*, **51**, 2528–2537.
10. Huey,R., Morris,G.M., Olson,A.J. and Goodsell,D.S. (2007) A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.*, **28**, 1145–1152.
11. Bayly,C.I., Cieplak,P., Cornell,W.D. and Kollman,P.A. (1993) A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges—the Resp model. *J. Phys. Chem.*, **97**, 10269–10280.
12. Jakalian,A., Bush,B.L., Jack,D.B. and Bayly,C.I. (2000) Fast, efficient generation of high-quality atomic Charges. AM1-BCC model: I. Method. *J. Comput. Chem.*, **21**, 132–146.
13. Rousseeuw,P.J. and Van Driessen,K. (2006) Computing LTS regression for large data sets. *Data Min. Knowledge Discov.*, **12**, 29–45.
14. Cheng,T.J., Li,X., Li,Y., Liu,Z.H. and Wang,R.X. (2009) Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf. Model.*, **49**, 1079–1093.
15. Wang,R.X., Lu,Y.P. and Wang,S.M. (2003) Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem*, **46**, 2287–2303.
16. Chang,D.T.H., Oyang,Y.J. and Lin,J.H. (2005) MEDock: a web server for efficient prediction of ligand binding sites based on a novel optimization algorithm. *Nucleic Acids Res.*, **33**, W233–W238.
17. Chang,D.T.H., Lin,J.H., Hsieh,C.H. and Oyang,Y.J. (2010) On the design of optimization algorithms for prediction of molecular interactions. *Int. J. Artif. Intelligence Tools*, **19**, 267–280.
18. Vaque,M., Arola,A., Aliagas,C. and Pujadas,G. (2006) BDT: an easy-to-use front-end application for automation of massive docking tasks and complex docking strategies with AutoDock. *Bioinformatics*, **22**, 1803–1804.
19. Cheng,A.C., Coleman,R.G., Smyth,K.T., Cao,Q., Soulard,P., Caffrey,D.R., Salzberg,A.C. and Huang,E.S. (2007) Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.*, **25**, 71–75.
20. Kellenberger,E., Muller,P., Schalon,C., Bret,G., Foata,N. and Rognan,D. (2006) sc-PDB: an annotated database of druggable binding sites from the protein data bank. *J. Chem. Inf. Model.*, **46**, 717–727.
21. Meslamani,J., Rognan,D. and Kellenberger,E. (2011) sc-PDB: a database for identifying variations and multiplicity of 'druggable' binding sites in proteins. *Bioinformatics*, **27**, 1324–1326.
22. Huang,Y., Niu,B.F., Gao,Y., Fu,L.M. and Li,W.Z. (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*, **26**, 680–682.
23. Shulman-Peleg,A., Nussinov,R. and Wolfson,H.J. (2005) SiteEngines: recognition and comparison of binding sites and protein–protein interfaces. *Nucleic Acids Res.*, **33**, W337–W341.
24. Shindyalov,I.N. and Bourne,P.E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.*, **11**, 739–747.
25. O'Boyle,N.M., Banck,M., James,C.A., Morley,C., Vandermeersch,T. and Hutchison,G.R. (2011) Open Babel: an open chemical toolbox. *J. Cheminform*, **3**, 33.
26. Surleraux,D.L.N.G., Tahri,A., Verschueren,W.G., Pille,G.M.E., de Kock,H.A., Jonckers,T.H.M., Peeters,A., De Meyer,S., Azijn,H., Pauwels,R. *et al.* (2005) Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor. *J. Med. Chem*, **48**, 1813–1822.
27. Zahler,S., Tietze,S., Totzke,F., Kubbutat,M., Meijer,L., Vollmar,A.M. and Apostolakis,J. (2007) Inverse *in silico* screening

for identification of kinase inhibitor targets. *Chem. Biol.*, **14**, 1207–1214.

28. Liu,F.L., Kovalevsky,A.Y., Tie,Y.F., Ghosh,A.K., Harrison,R.W. and Weber,I.T. (2008) Effect of flap mutations on structure of HIV-1 protease and inhibition by saquinavir and darunavir. *J. Mol. Biol.*, **381**, 102–115.

29. Kovalevsky,A.Y., Tie,Y.F., Liu,F.L., Boross,P.I., Wang,Y.F., Leshchenko,S., Ghosh,A.K., Harrison,R.W. and Weber,I.T. (2006) Effectiveness of nonpeptide clinical inhibitor TMC-114 on HIV-1 protease with highly drug resistant mutations D30N, I50V, and L90M. *J. Med. Chem*, **49**, 1379–1387.

30. Tie,Y.F., Boross,P.I., Wang,Y.F., Gaddis,L., Hussain,A.K., Leshchenko,S., Ghoshl,A.K., Louis,J.M., Harrison,R.W. and Weber,I.T. (2004) High resolution crystal structures of HIV-1 protease with a potent non-peptide inhibitor (UIC-94017) active against multi-drug-resistant clinical strains. *J. Mol. Biol.*, **338**, 341–352.

31. Kosmopoulou,M.N., Leonidas,D.D., Chrysina,E.D., Bischler,N., Eisenbrand,G., Sakarellos,C.E., Pauptit,R. and Oikonomakos,N.G. (2004) Binding of the potential antitumour agent indirubin-5-sulphonate at the inhibitor site of rabbit muscle glycogen phosphorylase b—comparison with ligand binding to pCDK2-cyclin A complex. *Eur J. Biochem.*, **271**, 2280–2290.

32. Sing,T., Sander,O., Beerenwinkel,N. and Lengauer,T. (2005) ROCR: visualizing classifier performance in R. *Bioinformatics*, **21**, 3940–3941.

33. Bressi,J.C., Jennings,A.J., Skene,R., Wu,Y.Q., Melkus,R., De Jong,R., O'Connell,S., Grimshaw,C.E., Navre,M. and Gangloff,A.R. (2010) Exploration of the HDAC2 foot pocket: synthesis and SAR of substituted *N*-(2-aminophenyl)benzamides. *Bioorg. Med. Chem. Lett*, **20**, 3142–3145.

34. Gruber,V.A., Rainey,P.M., Moody,D.E., Morse,G.D., Ma,Q., Prathikanti,S., Pade,P.A., Alvanzo,A.A.H. and McCance-Katz,E.F. (2012) Interactions between buprenorphine and the protease inhibitors darunavir–ritonavir and fosamprenavir–ritonavir. *Clin. Infect. Dis.*, **54**, 414–423.

35. Brown,K.C., Paul,S. and Kashuba,A.D.M. (2009) Drug interactions with new and investigational antiretrovirals. *Clin. Pharmacokinet.*, **48**, 211–241.

36. van de Waterbeemd,H. and Gifford,E. (2003) ADMET in silico modelling: towards prediction paradise? *Nat. Rev. Drug Discov.*, **2**, 192–204.

37. Cruciani,G., Carosati,E., De Boeck,B., Ethirajulu,K., Mackie,C., Howe,T. and Vianello,R. (2005) MetaSite: understanding metabolism in human cytochromes from the perspective of the chemist. *J. Med. Chem*, **48**, 6970–6979.

38. Aller,S.G., Yu,J., Ward,A., Weng,Y., Chittaboina,S., Zhuo,R.P., Harrell,P.M., Trinh,Y.T., Zhang,Q.H., Urbatsch,I.L. *et al.* (2009) Structure of *P*-glycoprotein reveals a molecular basis for poly-specific drug binding. *Science*, **323**, 1718–1722.